

Bootstrap in generalisierten linearen Modellen

Rothe, Günter

Veröffentlichungsversion / Published Version

Forschungsbericht / research report

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Rothe, G. (1986). *Bootstrap in generalisierten linearen Modellen*. (ZUMA-Arbeitsbericht, 1986/11). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-66409>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Günter Rothe

**BOOTSTRAP
IN
GENERALISIERTEN LINEAREN MODELLEN**

ZUMA-Arbeitsbericht Nr. 86/11

Summary

The theory of *Generalized Linear Models* allows the consideration of a number of statistical models under a unified approach. Usually (as in GLIM, e.g.), the unknown parameters are calculated using the maximum likelihood method. However, exact distributions of the estimates or of the goodness-of-fit measures (as e.g. the deviance) in general are unknown, hence exact confidence limit or tests do not exist. Asymptotic properties have been considered, among others, by Fahrmeir and Kaufmann (1985).

The Bootstrap technique (Efron, 1979) has been shown to be a good alternative to classical asymptotics in many cases. So it might be assumed that this remains true also in the case of generalized linear models. The present note illustrates the application of this technique and gives some theoretical properties of the bootstrap estimator.

1. Einleitung. Die Theorie der Generalisierten Linearen Modelle (Nelder & Wedderburn, 1972) ermöglicht es, eine große Klasse statistischer Modelle unter einem einheitlichen Ansatz darzustellen. Zu dieser Klasse gehören neben dem Klassischen Linearen Modell auch die log-linearen Modelle sowie die Modelle der logistischen Regression, die als Verfahren zur Auswertung qualitativer Daten in den letzten Jahren zunehmend an Beliebtheit gewonnen haben. Die Schätzung der in diesen Modellen interessierenden Parameter erfolgt in der Regel unter Verwendung des Maximum-Likelihood-Prinzips. Diese liefert jedoch nur Punktschätzer, d.h. exakte Verteilungen der Schätzer oder geeigneter Standardisierungen – wie etwa im Klassischen Linearen Modell unter Verwendung der t - bzw. der F -Verteilung – sind allgemein nicht explizit möglich, somit steht auch keine Methode zur exakten Genauigkeitsangabe für die geschätzten Parameter in Form von Konfidenzbereichen oder eine Methode zum Testen statistischer Hypothesen über diese Parameter zur Verfügung. Man nutzt daher oft Aussagen über das asymptotische Verhalten der Verteilung von Maximum-Likelihood-Schätzern, um sich so zumindest bei großen Stichproben ein Bild von der Genauigkeit des Schätzers zu verschaffen. Haberman (1977, 1984) sowie Fahrmeir & Kaufmann (1985) haben sich mit der Asymptotik solcher Schätzer auseinandergesetzt. Es besteht jedoch – wie in der Regel bei asymptotischen Aussagen – nach wie vor das Problem, daß die Asymptotik in konkreten Fällen häufig einfach

deshalb noch nicht greift, weil die Zahl der Beobachtungen für die Rechtfertigung asymptotischer Verfahren noch zu klein ist: Haben etwa die untersuchten qualitativen Variablen viele Ausprägungen, so werden oft die Besetzungszahlen der einzelnen Zellen so gering sein, daß die Verwendung der Asymptotik zu nicht mehr akzeptablen Schlußfolgerungen führt. Dies gilt insbesondere auch bei der Durchführung statistischer Tests bzw. beim Vergleich verschiedener Modellansätze unter Verwendung diverser Anpassungsmaße (im Generalisierten Linearen Modell etwa der *deviance*). Es besteht somit der Bedarf an Methoden, die in der Lage sind, die Verteilung der Schätzer (bzw. geeigneter Standardisierungen hiervon) besser zu approximieren, als dies unter Verwendung asymptotischer Verteilungen möglich ist. Hier bietet sich unter anderem das Bootstrap-Verfahren an, das sich in vielen (anderen) Fällen als eine solche Verbesserungsmöglichkeit herausgestellt hat.

Das Bootstrap-Verfahren wurde von Efron (1979) ursprünglich als Verfahren vorgeschlagen, um naheliegende Ad-hoc-Schätzer, die sich aufgrund der fehlenden Möglichkeit einer geschlossenen Darstellung nicht explizit berechnen lassen, durch eine Monte-Carlo-Studie zu schätzen bzw. durch geschickte Taylor-Entwicklung zu approximieren (vgl. auch Efron (1982), Efron und Tibshirani (1986)). Untersuchungen von Singh (1981), Bickel & Freedman (1981) sowie Beran (1982, 1984) haben gezeigt, daß Bootstrap-Schätzer als Schätzer für die Verteilung von Schätzern oft bessere Approximationen an die wahre Verteilung darstellen als deren

Approximation durch ihre asymptotische Verteilung. Dies legt die Vermutung nahe, daß ein ähnliches Verhalten auch im Bereich der Generalisierten Linearen Modelle zu finden ist. Eine Mindestvoraussetzung hierfür ist allerdings die Konsistenz der jeweiligen Bootstrap-Schätzer, d.h. die durch Bootstrap gewonnene Schätzung für die Verteilung des Schätzers sollte fast sicher oder zumindest nach Wahrscheinlichkeit die gleiche asymptotische Verteilung besitzen wie die tatsächliche Grenzverteilung des Schätzers (Eine genauere Spezifikation des Konsistenzbegriffs folgt). Sauermann (1986) hat dies für den Spezialfall der log-linearen Modelle nachgewiesen, sein Ergebnis geht sogar noch weiter: Indem zusätzlich mit wachsendem Stichprobenumfang sogar eine wachsende Zahl von Zellen und damit Parametern zugelassen ist, ist eine weitergehende Rechtfertigung der Verwendung des Bootstrap-Verfahren in Situationen mit geringen Zellbesetzungszahlen gegeben. Simulationsstudien bestärken die Vermutung.

In der vorliegenden Notiz zeigen wir, daß diese Konsistenz im Generalisierten Linearen Modell speziell im Falle natürlicher Link-Funktionen in der Tat stets gegeben ist, sobald auch die von Fahrmeir & Kaufmann (1985) formulierten Voraussetzungen für die asymptotische Normalität des ML-Schätzers erfüllt sind. Es ist zu vermuten, daß eine entsprechende Aussage auch für nicht-natürliche Link-Funktionen nachzuweisen ist, eine Untersuchung hierüber ist geplant. Unser Ergebnis umfaßt das von Sauermann (1986) somit

zunächst nur für den Fall einer Poisson-verteilten Kontingenztafel und fester (also von Stichprobenumfang unabhängiger) Dimension des Parameterraumes.

Es muß an dieser Stelle darauf hingewiesen werden, daß der in dieser Arbeit durchgeführte Konsistenznachweis nur sicherstellt, daß die Approximation an die Verteilung des (standardisierten) Schätzers durch das Bootstrap-Verfahren mindestens ebenso gut ist wie durch die asymptotische Verteilung. Es kann daraus noch nicht abgeleitet werden, daß durch Bootstrap tatsächlich eine Verbesserung des oben angesprochenen Dilemmas bei kleinen Stichprobenumfängen erreicht werden kann. Das Ergebnis ist als ein erster Schritt zu sehen, weiter Untersuchungen sind erforderlich, hierzu gehören Simulationsstudien und die Analyse asymptotischer Entwicklungen.

Im vorliegenden Beitrag wird zunächst die im folgenden verwendete Terminologie im Zusammenhang mit dem Generalisierten Linearen Modell eingeführt, danach wird in Abschnitt 3 das Bootstrap-Verfahren allgemein kurz skizziert und für den Fall des Generalisierten Linearen Modells präzisiert sowie schließlich die angekündigten Konsistenzaussagen formuliert; Abschnitt 4 geht dann noch auf ein etwas allgemeineres Modell ein. Die erforderlichen mathematischen Beweise finden sich im Anhang. Im Abschnitt 5 schließlich wird anhand eines Datensatzes die Durchführung der Bootstrap-Schätzprozedur skizziert; die Arbeit schließt mit einer

an diesen Datensatz angelehnten Simulationsstudie, die einen Einblick in die Wirkungsweise des Verfahrens liefert. Die Bootstrap-Verfahren in diesem Abschnitt wurden mit der Programmiersprache GLIM realisiert, bei den verwendeten Programm handelt es sich allerdings derzeit noch um eine Ad-hoc-Prozedur. Eine Verfeinerung, die einen komfortableren Zugriff ermöglicht, ist in Vorbereitung.

2. Generalisierte Lineare Modelle. Wir werden uns in dieser Arbeit vorwiegend sehr eng an die bei Fahrmeir & Kaufmann (1985) verwendete Terminologie halten und fassen im folgenden nur kurz die wesentlichsten dort präzisierten Bezeichnungen und Regularitätsbedingungen zusammen. Die Arbeit werden wir stets kurz mit F&K referieren.

\mathbf{y}_n , $n \in \mathbb{N}$, seien unabhängige Zufallsgrößen mit Dichten

$$f(\mathbf{y}_n | \theta_n) = c(\mathbf{y}_n) \exp(\theta_n' \mathbf{y}_n - b(\theta_n)), \quad n=1,2,\dots$$

Hierbei sei jeweils $\theta_n = u(\gamma_n)$ für eine bekannte Funktion u und

$$\gamma_n = \mathbf{z}_n' \beta$$

wobei der Regressor \mathbf{z}_n bekannt und $\beta \in \mathbb{R}^p$ der zu schätzende unbekannte Parameter ist.

In diesem Fall gehorchen die Zufallsgrößen Y_i , $i \in N$ einem Generalisierten Linearen Modell .

Im folgenden werden wir uns ausschließlich mit dem Spezialfall $u=id$ befassen. Dann lautet die Score-Funktion

$$\begin{aligned} s_n(\beta) &= \partial(\sum_{i=1}^n \log(f(Y_i | Z_i' \beta))) / \partial \beta \\ &= \sum_{i=1}^n \partial l_i(\beta) / \partial \beta = \sum_{i=1}^n Z_i (Y_i - \mu_i(\beta)) \end{aligned}$$

und die Informationsmatrix

$$F_n(\beta) = \sum_{i=1}^n Z_i X_i(\beta) Z_i' .$$

Hierbei ist

$$\mu_i(\beta) = \mu(u(Y_i)) = \mu(\theta_i) = E_{\beta_i}(Y_i) = b'(\theta_i)$$

und

$$X_i(\beta) = \text{Kov}_{\beta}(Y_i) = -\partial^2 l_i(t) / \partial t \partial t' \Big|_{t=Z_i' \beta} .$$

Ferner setzen wir die gleichen Regularitätsbedingungen wie in Abschnitt 2.2 von F&K voraus und verwenden die Bezeichnungen $A^{1/2}$ bzw. $A^{T/2}$ analog zum Abschnitt 3.1 dort.

Mit $\lambda_{\min}(A)$ wird der kleinste Eigenwert der Matrix A bezeichnet.
Ergänzend zu F&K bezeichnen wir für $\beta_0 \in B$, $\delta > 0$ und $n=1,2,\dots$

$$N_n(\delta, \beta_0) = \{ \beta : \left\| F_n^{T/2}(\beta - \beta_0) \right\| \leq \delta \}.$$

und

$$W_n(\delta, \beta_0) = \max_{\beta \in N_n(\delta, \beta_0)} \left\| V_n(\beta) - I \right\|$$

wobei $V_n(\beta) = (F_n(\beta_0))^{-1/2} F_n(\beta) (F_n(\beta_0))^{-T/2}$

Dann gilt

Satz 1 (vgl F&K, Thms. 1 und 3)

Es gelte

$$(N) \quad W_n(\delta, \beta_0) \rightarrow 0 \quad \text{für alle } \delta > 0.$$

sowie

$$(D) \quad \lambda_{\min}(F_n(\beta_0)) \rightarrow \infty$$

Dann existiert eine Folge $\{\hat{\beta}_n\}$ von Zufallsgrößen derart, daß

$$(1) \quad P_{\beta_0} (S_n(\hat{\beta}_n) = 0) \rightarrow 1$$

$$(ii) \quad \hat{\beta}_n \rightarrow \beta_0 \text{ nach } P_{\beta_0} \text{-Wahrscheinlichkeit}$$

$$(iii) \quad D_{\beta_0} (F_n(\beta_0))^{T/2} (\hat{\beta}_n - \beta_0) \rightarrow N(0, I).$$

Wir schließen diesen Abschnitt mit einigen Bemerkungen:

1. Da F_n positiv definit ist, ergibt sich, daß die Lösung $\hat{\beta}_n$ von $s_n(\hat{\beta}_n)=0$ eindeutig ist, sofern sie überhaupt existiert.
2. Für jedes $\beta_0 \in B$ und jedes $\delta > 0$ gilt, wie man leicht sieht,

$$W_n(\delta, \beta_0) = \sup_{0 \neq \tau \in R^q} \sup_{\beta \in N_n(\delta, \beta_0)} \left| \frac{\tau' F_n(\beta) \tau}{\tau' F_n(\beta_0) \tau} - 1 \right|$$

Diese Darstellungen werden wir in Abschnitt 3 benötigen.

3. In F&K ist eine Vielzahl von Voraussetzungen und Spezialfällen angegeben, unter denen die Bedingungen (N) und (D) erfüllt sind. Diese Voraussetzungen sind dann nicht abhängig von Parameter β und decken nahezu alle praktisch auftretenden Fälle tatsächlich ab, insbesondere wird dort auch der Fall stochastischer Regressoren berücksichtigt. Die Bedingungen (D) und (N) sind also keine sehr restriktive und somit "esotherische" Forderung.

3. Die Bootstrap-Verteilung. Im soeben definierten Generalisierten Linearen Modell hängt die gemeinsame Verteilung der Zufallsgrößen Y_1, \dots, Y_n nur vom zugrundeliegenden wahren Parameter $\beta_0 \in B$ ab. Ist nun $S_n(Y_1, \dots, Y_n; \beta)$ eine R^k -wertige Funktion der Daten und des Parameters - derartige Funktionen werden wir im folgenden kurz als parametrische Statistiken bezeichnen -, so ist ihre Verteilung

$$D_n^{S(\beta)} = D_\beta(S_n(Y, \beta))$$

eine Funktion des unbekannten Parameters β .

Die Idee des Bootstrap-Verfahrens besteht nun einfach darin, die Größe $D_n^{S(\beta)}$ durch $D_n^{S(\hat{\beta}_n)}$ zu schätzen. Dies ist somit ein Schätzer, dessen Wertebereich nicht, wie meist üblich, eine Teilmenge eines endlich-dimensionalen Unterraumes über R ist, sondern die Menge der Wahrscheinlichkeitsverteilungen auf R^p . Wir wollen an dieser Stelle nicht auf die hierdurch entstehenden Meßbarkeitsprobleme eingehen und weisen nur darauf hin, daß sich nach Übergang von den Verteilungen zu den Verteilungsfunktionen der Bootstrap-Schätzer als stochastischer Prozeß mit p -dimensionalem "Zeit"-Parameter erweist.

Es muß hier darauf hingewiesen werden, daß die Definition des Bootstrap-Schätzers wesentlich einfacher aussieht als seine tatsächliche Berechnung schließlich wirklich ist. In der Regel wird

man die Bootstrap-Verteilung bzw. ihre Verteilungsfunktion mit einer Monte-Carlo-Studie schätzen müssen: Man gibt einen Monte-Carlo-Stichprobenumfang N vor, generiert N Zufallsvektoren $(Y_{1k}^*, \dots, Y_{nk}^*)$, $k=1, \dots, N$ und bestimmt jeweils die Werte der parametrischen Statistik $S_n(Y_{1k}^*, \dots, Y_{nk}^*; \hat{\beta}_n)$ für $k=1, \dots, N$; die aus diesen N Werten gebildete empirische Verteilungsfunktion bildet dann einen geeigneten Schätzer für den Bootstrap-Schätzer von $D_n^S(\beta)$. Wir werden diese zweite – bei konkreten Rechnungen allerdings praktisch erforderliche – Stufe bei der Bestimmung des Bootstrap-Schätzers jedoch zunächst ignorieren und uns nur mit Eigenschaften der Größe $D_n^S(\hat{\beta}_n)$ auseinandersetzen.

Diesen Bootstrap-Schätzer bezeichnen wir nun als *schwach* (bzw. *streng*) *d-konsistent* in $\beta \in B$, falls

$$d(D_n^S(\hat{\beta}_n), D_n^S(\beta)) \rightarrow 0 \quad \text{nach } P_\beta\text{-Wahrscheinlichkeit} \\ \text{(bzw. } P_\beta\text{-fast sicher).}$$

Hierbei bezeichne d eine geeignet gewählte Metrik auf der Menge der Verteilungen

$$\{ D_n^S(\beta), \beta \in B \}$$

der parametrischen Statistik $S_n(y; \beta)$. Im folgenden wird in dieser Arbeit speziell die Prohorov-Metrik d_p betrachtet, die die

schwache Topologie auf der Menge der Verteilungen über den Borel-schen Mengen des \mathbb{R}^k generiert. Besitzt nunmehr $D_n^S(\beta)$ in diesem Sinne eine Grenzverteilung (etwa D_∞), so bedeutet die strenge Konsistenz explizit, daß die Folge von ML-Schätzern $\{\hat{\beta}_n\}$ P_β -fast sicher eine Folge von Bootstrap-Schätzern $\{D_n^S(\hat{\beta}_n)\}$ produziert, die die gleiche Grenzverteilung D_∞ besitzt. Konsistenz nach Wahrscheinlichkeit liegt nach einem elementaren Satz aus der Wahrscheinlichkeitsrechnung dann bekanntlich vor, wenn zumindest jede Teilfolge von $\{D_n^S(\hat{\beta}_n)\}$ eine ihrerseits fast sicher gegen D_∞ konvergente Teilfolge besitzt. Beabsichtigt man darüberhinaus jedoch etwa, auch die Konsistenz gewisser höherer Momente von $D_n^S(\hat{\beta}_n)$ sicherzustellen, so bietet sich etwa die Verwendung der Mallows-Distanz d_q für ein $q > 0$ an (vgl. hierzu etwa Bickel & Friedman 1981).

Im folgenden betrachten wir zunächst die schwache d_p -Konsistenz der Bootstrap-Verteilung des standardisierten Schätzfehlers

$$S_n(Y, \beta_0) = (F_n(\beta_0))^{-T/2} (\hat{\beta}_n - \beta_0):$$

Mit der Bezeichnung

$$D_n^S(\beta) = D_\beta(S_n(Y, \beta))$$

zeigen wir

Satz 2:

Unter (D) und (N) ist der Bootstrap-Schätzer $D_n^S(\hat{\beta}_n)$ von $D_n^S(\beta)$ schwach d_p -konsistent.

Der Beweis dieses Satzes wird explizit im Anhang durchgeführt.

In ihrer Arbeit weisen F&K darauf hin, daß bei der praktischen Anwendung ihres Theorems 3 (Satz 1 (ii) in der vorliegenden Arbeit) in der Regel $(F_n(\beta_0))^{T/2}$ durch einen Schätzer hierfür, also etwa $(F_n(\hat{\beta}_n))^{T/2}$ ersetzt werden muß. Dies sei aber nur möglich, wenn entweder die Bedingung (N) leicht verschärft wird ((Q) bei F&K) oder eine spezielle Quadratwurzel (nämlich die Cholesky-Wurzel) verwendet wird. Bei der Verwendung der Bootstrap-Verteilung ist diese Einschränkung nicht erforderlich.

Die Tatsache, daß zur Sicherung der Bootstrap-Konsistenz des standardisierten Schätzfehlers in Satz 2 keine weitergehenden Voraussetzungen als in Satz 1 erforderlich waren, erlaubt es nun, auch von der Bootstrap-Konsistenz ausgehen zu können, wenn die bei F&K im Abschnitt 3.3 angesprochenen Voraussetzungen in Spezialfällen untersucht werden müssen. Ein gewisses Problem tritt allerdings bei der Modellannahme stochastischer Regressoren auf: Führt man in diesem Falle "naiv" das Bootstrap-Verfahren so durch wie bisher, d.h. ignoriert man die "Zufälligkeit" der Größen Z_1, \dots, Z_n bei der Konstruktion des Bootstrap-Schätzers für $D_n^S(\beta)$.

so sind unter den Voraussetzungen (R_S) von F&K die Annahmen (D) und (N) fast sicher erfüllt und $D_n^S(\hat{\beta}_n)$ konvergiert in der Tat ebenfalls gegen $N(0, I)$. Allerdings verstößt man an dieser Stelle eigentlich gegen das Bootstrap-Prinzip, das ja darauf basiert, daß man zunächst die das Modell beschreibenden Parameter schätzt und die Verteilung der interessierenden parametrischen Statistik unter diesem Parameter als Schätzung für ihre tatsächliche Verteilung heranzieht. Geht man aber vom Modell mit stochastischen Regressoren aus, nimmt man also an, daß auch die Regressoren Realisationen unabhängiger, identisch verteilter Zufallsgrößen mit Verteilung D_Z darstellen, so ist diese Verteilung ein zusätzlicher unbekannter Parameter im Modell und muß, will man dem Bootstrap-Konzept strikt folgen, ebenfalls "sinnvoll" geschätzt werden. Hier würde sich nun etwa die empirische Verteilung anbieten, die wie im folgenden mit $\hat{D}_{Z,n}$ bezeichnen werden.

Tatsächlich ist im Modell der stochastischen Regressoren die Verteilung des standardisierten Schätzfehlers nicht nur eine Funktion von β , sondern auch von D_Z ; richtiger schreiben wir also hierfür $D_n^S(\beta, D_Z)$. Der dem Bootstrap-Prinzip folgende Schätzer hierfür wäre dann $D_n^S(\hat{\beta}_n, \hat{D}_{Z,n})$, und es ist zu erwarten, daß der die tatsächliche Verteilung besser approximiert als der oben angesprochene "naive" Schätzer, der sich in der hier verwendeten Terminologie beschreiben ließe als $D_n^S(\hat{\beta}_n, \delta_n)$, wobei δ_n das Einpunktmaß auf (Z_1, \dots, Z_n) bezeichne. In der vorliegenden Arbeit

werden wir uns mit dem Konsistenznachweis für diesen Fall allerdings nicht befassen und uns ausschließlich um die "naive" Version kümmern, also um den "bedingten" Bootstrap, der nur die bedingte Verteilung von $\hat{\beta}_n - \beta$ bei gegebenem (Z_1, \dots, Z_n) schätzt.

4. Generalisiertes Lineares Modell mit Skalenparameter. In der grundlegenden Arbeit über Generalisierte Lineare Modelle von Nelder & Wedderburn (1972) wird eigentlich eine etwas allgemeinere Modellklasse betrachtet: Die Autoren bezeichnen dort als Generalisiertes Lineares Modell eine Klasse mit Dichten

$$f(\mathbf{y}_n | \theta_n, \phi) = c(\mathbf{y}_n, \phi) \exp((\theta_n' \mathbf{y}_n - b(\theta_n))/a(\phi)), \quad n=1, 2, \dots$$

wobei $a(\phi)$ zusätzlich einen Skalenparameter darstellt. Die ML-Schätzung von β ist in diesem erweiterten Modell identisch mit der im bisher betrachteten.

Will man allerdings die Bootstrap-Schätzung der Verteilung des standardisierten Schätzfehlers untersuchen, so muß natürlich wiederum auch dieser Skalenparameter in die Modell- und danach in die Verteilungsschätzung eingehen. Wir wollen uns in diesem Abschnitt hiermit kurz auseinandersetzen:

Die Score-Funktion lautet in diesem allgemeineren Fall

$$\begin{aligned}\tilde{s}_n(\beta, \phi) &= \partial(\sum_{i=1}^n \log(f(\mathbf{y}_i | \mathbf{z}_i', \beta, \phi))) / \partial \beta \\ &= \sum_{i=1}^n \partial l_i(\beta, \phi) / \partial \beta = \sum_{i=1}^n \mathbf{z}_i (\mathbf{y}_i - \mu_i(\beta)) / a(\phi).\end{aligned}$$

Mit der Bezeichnung

$$\mathbf{I}(\theta) = \partial^2 b(\theta) / \partial \theta \partial \theta'$$

ergibt sich dann

$$\mathbf{I}(\theta, \phi) := \text{cov}_{\theta, \phi}(\mathbf{Y}_i) = a(\phi) \cdot \mathbf{I}(\theta).$$

und somit

$$\mathbf{F}_n(\beta, \phi) = \text{Kov}_{\beta, \phi}(\mathbf{s}_n(\beta, \phi)) = \sum_{i=1}^n \mathbf{z}_i \mathbf{I}(\theta) \mathbf{z}_i' / a(\phi) \quad .$$

Schließlich gilt

$$\begin{aligned}\partial^2 \left(\sum_{i=1}^n \log(f(\mathbf{y}_i | \mathbf{z}_i', \beta, \phi)) \right) / \partial \beta \partial \beta' &= - \sum_{i=1}^n \mathbf{z}_i \mathbf{I}(\theta) \mathbf{z}_i' / a(\phi) \\ &= \mathbf{F}_n(\beta, \phi).\end{aligned}$$

Verwenden wir nun noch die Bezeichnung

$$\mathbf{F}_n(\beta) = \sum_{i=1}^n \mathbf{z}_i \mathbf{I}(\theta) \mathbf{z}_i',$$

so ergibt sich folgende Aussage:

Satz 3:

Für festes (β_0, ϕ_0) gelte (D) und (N). Dann gilt für die parametrische Statistik

$$\tilde{S}_n(\mathbf{Y}; \beta, \phi) = (a(\phi_0) \cdot F_n(\beta_0))^{T/2} \cdot (\hat{\beta}_n - \beta_0)$$

die Aussage

$$D_n \tilde{S}(\beta_0, \phi_0) \rightarrow N(0, I) .$$

Ist zusätzlich der Maximum-Likelihood-Schätzer $\hat{\phi}_n$ von ϕ (schwach) konsistent und $a(\phi)$ stetig in ϕ_0 , so gilt

$$D_n \tilde{S}(\hat{\beta}_n, \hat{\phi}_n) \rightarrow N(0, I) .$$

für den Bootstrap-Schätzer von $D_n \tilde{S}(\beta_0, \phi_0)$.

Der Beweis dieser Aussage findet sich ebenfalls im Anhang.

5. Ein Beispiel. Im folgenden Abschnitt wollen wir anhand eines Datensatzes Probleme bei der Durchführung der Bootstrap-Prozedur beschreiben und die Ergebnisse diskutieren. Wir verwenden hierzu einen Auszug aus table 5.4 bei HABERMAN(1978): Im General Social Survey der Jahre 1974 und 1975 beantworteten 1305 Männer,

deren Bildungsniveau (in Form der Variablen *years of education*) bekannt war, die Frage, ob sie der folgenden Aussage eher zustimmen oder sie eher ablehnen würden: *Women should take care running their homes and leave running the country up to men.* Die Antwortverteilung ist in Tab.1 wiedergegeben. Diesen Daten wird nun das Modell zugrundegelegt, daß jeweils $p(t)$ die Wahrscheinlichkeit angibt, daß eine Person mit Bildungsniveau t (= *years of education*) der Aussage zustimmt und der Zusammenhang zwischen den unbekannten Parametern und $p(t)$ durch die Beziehung

$$\text{logit } p(t) = \beta_0 + \beta_1 t$$

gegeben ist, d.h. die Beobachtungen $y_i, i=1, \dots, 1305$ sind unabhängige Bernoulli-verteilte Zufallsvariablen mit

$$P(y_i=1)=p_i$$

mit

$$\text{logit } p_i = \beta_0 + \beta_1 t_i ,$$

und es ist

$$z_i = (1, t_i)' ; \quad \beta = (\beta_0, \beta_1)' .$$

Unter Verwendung des Programmpaketes GLIM 3.77 (vgl. PAYNE (1985)) ergab sich als Schätzer für β

$$(\hat{\beta}_0, \hat{\beta}_1)' = (2.089, -0.0234)'$$

mit einer Kovarianzmatrix von

$$\hat{\text{Kov}}(\hat{\beta}) = \begin{bmatrix} 0.0555 & 0.0046 \\ 0.0046 & 0.000407 \end{bmatrix}$$

Für β_1 ergäbe sich also aufgrund der Normalapproximation bei einer geschätzten Standardabweichung von $s_1=0.02019$ ein zweiseitiges (asymptotisches) 95%-Konfidenzintervall von

$$\begin{aligned} I(\beta_1) &= [\hat{\beta}_1 - 1.96 s_1, \hat{\beta}_1 + 1.96 s_1] \\ &= [-0.2736, -0.1944] \end{aligned}$$

Die Rechtfertigung für die Verwendung der Normalapproximation ergibt sich aus der Anwendung etwa von Satz 1. Soll nun stattdessen ein Konfidenzintervall für β_1 unter Verwendung des Bootstrap-Prinzips konstruiert werden, so bedeutet das nun in unserem konkreten Beispiel, daß zunächst die Verteilungsfunktion der Zufallsgröße $\hat{\beta}_1 - \beta_1$ für den Fall berechnet werden müßte, daß der tatsächlich zugrundeliegende, das Modell beschreibende Parametervektor genau die Größe

$$\beta = (2.089, -0.0234)'$$

ist. Praktisch ist dies jedoch nicht durchführbar; aus diesem Grunde wird diese Verteilungsfunktion durch eine Monte-Carlo-Studie (deren Umfang im folgenden mit N_{boot} bezeichnet wird) geschätzt: In jedem einzelnen MC-Schritt werden somit - unter Verwendung eines (Pseudo-)Zufallszahlengenerators - Bernoulli-verteilte Zufallsgrößen y_i^* , $i=1, \dots, 1305$ mit

$$\text{logit } P(y_i^*=1) = 2.089 - 0.0234 t_i$$

generiert, anschließend wird dann jeweils der aus diesen Daten resultierende Schätzer $\hat{\beta}_1^*$ berechnet und gespeichert. Die aus diesen N_{boot} Werten gebildete empirische Verteilungsfunktion fungiert dann als ein Schätzer für die eigentliche Bootstrap-Verteilung von $\hat{\beta}_1$. Konkret wurde diese MC-Studie -ebenfalls unter der Verwendung von GLIM 3.77 - mit $N_{\text{boot}}=1000$ durchgeführt. Als untere bzw. obere Grenze für ein Bootstrap-Konfidenzintervall ist in diesem Fall das 25-kleinste bzw. 25-größte $\hat{\beta}_1^*$ zu wählen. Es ergab sich

$$I^*(\beta_1) = (-0.2759, -0.1986]$$

Man sieht also, daß sich die beiden Konfidenzintervalle I und I^* eigentlich nur unwesentlich unterscheiden. Dies läßt sich natürlich direkt mit der Normalapproximation erklären: Der vorliegende

Stichprobenumfang ist sicherlich hinreichend groß, um die Normalapproximation $I(\beta_1)$ zu rechtfertigen, das gleiche gilt natürlich auch für $I^*(\beta_1)$. Hier rentiert sich also der große Rechenaufwand des Bootstrap-Verfahrens (er ist ja praktisch N_{boot} -mal so groß) offenbar nicht.

Um zu demonstrieren, daß die Situation bei kleineren Stichprobenumfängen etwas anders gelagert ist, haben wir nun zufällig eine Stichprobe vom Umfang 100 aus den oben beschriebenen Befragten gezogen. Die auf diese Weise gewonnene Teilstichprobe ist in Tab. 2 angegeben. Gehen wir nun davon aus, daß ausschließlich diese Beobachtungen zur Verfügung stehen, so können analog zum oben beschriebenen Vorgehen nun wieder β und die Bootstrap-Verteilung von $\hat{\beta}_1$ bestimmt werden. Es ergeben sich in diesem Falle folgende Schätzwerte:

$$\hat{\beta} = (1.854, -0.2455)'$$

$$\hat{s}_1 = 0.08176$$

Das Konfidenzintervall unter Verwendung der Normalapproximation ist somit

$$I_0(\beta_1) = [-0.4058, -0.0852].$$

Die Bootstrap-Prozedur hingegen - wieder mit einer MC-Studie vom Umfang $N_{\text{boot}}=1000$ - liefert den Konfidenzbereich

$$I_0^*(\beta_1) = [-0.4578, -0.1092].$$

Dieses Intervall ist also etwas größer als I_0 und außerdem etwas nach links verschoben. Bei diesem (kleinen) Stichprobenumfang sind also das durch Normalapproximation und das durch Bootstrap gewonnene Intervall deutlich verschieden. Beide halten das Niveau 95% nicht exakt, sondern (aufgrund von Satz 1 und Satz 2) nur asymptotisch ein, es kann aufgrund der bisher durchgeführten Untersuchungen also nicht festgestellt werden, welches der beiden Verfahren die tatsächlich zugrundeliegende Verteilung besser approximiert und damit genauere Konfidenzbereiche liefert. Dies liegt natürlich daran, daß der wahre Parameter β unbekannt ist und so die Schätzung mit der "Realität" nicht tatsächlich verglichen werden kann. Zur Überprüfung des Verhaltens der Bootstrap-Verteilung wäre also eine Simulationsstudie erforderlich, in der ein künstliches Modell konstruiert wird.

Wir wollen dies im folgenden durchführen, wobei wir uns bei der Wahl des simulierten Modells möglichst eng an das oben beschriebene Beispiel anlehnen. Wir betrachten $n=100$ Beobachtungseinheiten, die Einflußgröße t habe die gleiche Struktur wie in Tab. 2 und der die Wahrscheinlichkeiten

$$P(Y_1=1)=p_1 \quad \text{mit} \quad \text{logit } p_1 = \beta_0 + \beta_1 t_1$$

generierende Parametervektor $\beta=(\beta_0, \beta_1)$ sei gegeben durch

$$\beta = (1.854, -0.2455)'$$

Die Bestimmung der Verteilung von $\hat{\beta}_1$ unter dieser Parameterkonstellation kann nun durch eine MC-Studie erfolgen. Im vorliegenden Fall ist dies allerdings bereits geschehen, nämlich bei der Berechnung des Bootstrap-Konfidenzintervalls I_0^* im obigen Beispiel. Eine komplette Simulationsstudie müßte nun folgendermaßen aussehen: Unter Verwendung der vorgegebenen Parameter wird ein Datensatz generiert und - ausgehend von diesem Datensatz - nun die Bootstrap-Schätzung der Verteilung von $\hat{\beta}_1 - \beta_1$ bzw. die beiden Konfidenzintervalle $I_0(\beta_1)$ und $I_0^*(\beta_1)$ bestimmt. Im nächsten Simulationsschritt wird nun ein weiterer Datensatz generiert etc.; als Schätzung für das tatsächliche Niveau der beiden Intervalle ist dann die relative Häufigkeit der Simulationsschritte zu sehen, in denen das Intervall den wahren Parameter β_1 tatsächlich enthält. Um hierfür eine hinreichend genaue Schätzung zu gewinnen (etwa eine Standardabweichung unter 0.01), sind allerdings viele Simulationsschritte erforderlich (über 1500); da Bootstrap seinerseits ebenfalls eine Monte-Carlo-Studie erfordert, sind insgesamt sehr viele Parameterschätzungen durchzuführen (also etwa $1500 \times N_{\text{boot}}$), was insgesamt ausgesprochen zeitaufwendig ist.

Aus diesem Grunde wurden - wie etwa auch bei Sauermann (1986) - nur drei Simulationsdurchgänge gestartet und in jedem Durchgang die tatsächliche Verteilung von $\hat{\beta}_1 - \beta_1$, ihre Normalapproximation $N(0, \hat{\sigma}_1^2)$ und ihre Bootstrap-Approximation verglichen. Die drei

Simulationsdurchgänge lieferten die folgenden Schätzungen für β und die daraus abgeleiteten Schätzer s_1 für die Varianz von $\hat{\beta}_1$:

$$(1) \hat{\beta} = (2.244, -0.2757)' , \quad s_1 = 0.08447$$

$$(2) \hat{\beta} = (2.521, -0.3313)' , \quad s_1 = 0.09299$$

$$(3) \hat{\beta} = (0.924, -0.1278)' , \quad s_1 = 0.06990$$

Der erste der drei Simulationsdurchgänge produzierte also eine relativ genaue Schätzung des wahren Parameters, während die Schätzwerte für β_1 in Durchgang 2 und 3 dagegen relativ niedrig bzw. hoch liegen. In Fig.1 a-c lassen sich nun die jeweiligen Verteilungsfunktionen miteinander vergleichen. In jedem Fall stellt die durchgezogene Linie die (Monte-Carlo-Schätzung der) Verteilungsfunktion von $(\hat{\beta}_1 - \beta_1)$ dar, die gestrichelte Linie ist die Bootstrap-Schätzung dieser Verteilung und die dritte (Strichpunkt-)Kurve stellt die jeweilige Normalapproximation, also die Verteilungsfunktion von $N(0, s_1^2)$, dar. Natürlich ist im ersten Fall die Bootstrap-Schätzung der Normalapproximation überlegen. Aber auch in den beiden anderen Fällen, in denen die Schätzung von β_1 jeweils etwa eine Standardabweichung unter bzw. über dem wahren Parameter liegt, kann die Bootstrap-Schätzung nicht als schlechter bezeichnet werden.

Wenn auch der Vergleich der Schätzungen in drei Fällen nur wenig globale Aussagekraft besitzt, so sind die vorliegenden Untersuchungen durchaus ein Indiz dafür, daß auch im Fall Generalisierter Linearer Modelle das Bootstrap-Konzept sinnvoll sein kann.

Anhang: Beweise der Sätze 2 und 3. In diesem Abschnitt führen wir nun explizit die Konsistenzbeweise durch; zunächst weisen wir Satz 2 nach.

Wir skizzieren hierzu zunächst die weitere Vorgehensweise:

Sei $\tilde{B} = \{ \beta_n^* \}$ eine beliebige Folge in B mit $\beta_n^* \rightarrow \beta_0$,

$$\tilde{V}_n(\beta) = (F_n(\beta_n^*))^{-1/2} F_n(\beta) (F_n(\beta_n^*))^{-T/2},$$

sowie (vgl. Abschn. 1)

$$\tilde{W}_n(\delta) = W_n(\delta, \beta_n^*) = \sup_{\beta \in N_n(\delta, \beta_n^*)} \left\| \tilde{V}_n(\beta) - I \right\|.$$

Mit diesen Bezeichnungen möge \tilde{B} folgende Beziehungen erfüllen:

$$[D(\tilde{B})] \quad \lambda_{\min}(F_n(\beta_n^*)) \rightarrow \infty,$$

$$[N(\tilde{B})] \quad \tilde{W}_n(\delta) \rightarrow 0 \text{ für alle } \delta > 0.$$

Wir werden zunächst zeigen, daß diese Eigenschaften die Beziehung

$$D_{\beta_n^*}^* \left((F_n(\beta_n^*))^{T/2} (\hat{\beta}_n - \beta_n^*) \right) \rightarrow N(0, I)$$

implizieren. Im zweiten Schritt zeigen wir dann, daß im Fall

$$\{ \hat{\beta}_n^* \} = \{ \hat{\beta}_n \}$$

die Konvergenzaussagen $D(\hat{B}^*)$ und $N(\hat{B}^*)$ nach P_{β} -Wahrscheinlichkeit erfüllt sind. Die bereits angesprochene Teil-Teilfolgenargumentation liefert dann die Behauptung.

Zur Vorbereitung des Beweises von Satz 2 benötigen wir drei Lemmata. Zunächst betrachten wir die fest vorgegebene Folge $\hat{B}^* = \{\hat{\beta}_n^*\}$ in B . Ferner schreiben wir abkürzend \hat{F}_n^* bzw. \hat{s}_n^* für $F_n(\hat{\beta}_n^*)$ bzw. $s_n(\hat{\beta}_n^*)$. Die Aussagen des folgenden Lemmas sind Modifikationen von Aussage (3.8) sowie Lemma 1 und Theorem 3 bei F&K.

Lemma 1

Es seien $[D(\hat{B}^)]$ und $[N(\hat{B}^*)]$ erfüllt. Dann gilt:*

$$(A.1) \quad P_{\hat{\beta}_n}^* \left(\left\| \hat{F}_n^{*T/2} (\hat{\beta}_n - \hat{\beta}_n^*) \right\| \leq \delta \right) \geq 1 - \eta$$

für alle $n > n_1$.

$$(A.2) \quad D_{\hat{\beta}_n}^* (\hat{F}_n^{*-1/2} s_n) \rightarrow N(0, I).$$

$$(A.3) \quad D_{\hat{\beta}_n}^* \left(\hat{F}_n^{*T/2} (\hat{\beta}_n - \hat{\beta}_n^*) \right) \rightarrow N(0, I).$$

Beweis: Die Beweise in F&K, Th.1, L.1, Th.3 können im wesentlichen wörtlich übernommen werden, sofern man β_0 , $N_n(\delta)$, F_n , s_n ,

sowie V_n ersetzt durch β_n^* , $N_n(\delta, \beta_n^*)$, F_n^* , s_n^* , sowie \bar{V}_n . Ein Punkt ist jedoch hierbei zu beachten: Der einzige Grund, warum im Nachweis der Aussagen von Satz 1 (i) und (iii) bei F&K die Beziehung (D) benötigt wird, liegt darin, daß durch (D) folgende Aussage sichergestellt wird:

Zu $\delta > 0$ gibt es $n_1(\delta)$ derart, daß

$\partial N_n(\delta, \beta_0)$ für $n > n_1(\delta)$ vollständig in B enthalten ist.

$D(\bar{B})$ liefert nun die analoge Aussage für $\partial N_n(\delta, \beta_n^*)$. \square

Bemerkung: Gilt $N(\bar{B})$ nur für eine Teilfolge $\{\bar{W}_{k(n)}^*(\delta)\}$, so bleibt Lemma 1 richtig, sofern auch in (A.1)–(A.3) jeweils stets n durch $k(n)$ ersetzt wird.

Die nächste Aussage ist ein Hilfssatz für Lemma 3:

Lemma 2:

Sei $\beta^* \in N_n(C, \beta_0)$ und $\beta \in N_n(\delta, \beta^*)$ für $C > 0$, $\delta > 0$, $\beta \in B$, $\beta^* \in B$, $\beta_0 \in B$. Dann gilt

$$\beta \in N_n(\bar{\delta}_n, \beta_0) \quad \text{für jedes } \bar{\delta}_n \geq \delta (1 - \bar{W}_n(C, \beta_0))^{-1/2} + C.$$

Beweis: Wir schreiben wieder kurz \bar{F}_n^* für $F_n(\beta^*)$. Dann gilt unter den Voraussetzungen des Lemmas:

$$\begin{aligned}
& \left\| F_n^{T/2} (\beta - \beta_0) \right\| \\
& \leq \left\| F_n^{T/2} (\beta - \beta^*) \right\| + \left\| F_n^{T/2} (\beta^* - \beta_0) \right\| \\
& \leq \left\| F_n^{T/2} \tilde{F}_n^{-T/2} \right\| \cdot \left\| \tilde{F}_n^{T/2} (\beta - \beta^*) \right\| + C
\end{aligned}$$

Hierbei ist

$$\begin{aligned}
\left\| F_n^{T/2} \tilde{F}_n^{-T/2} \right\|^2 &= \lambda_{\max}(F_n^{T/2} \tilde{F}_n^{-T/2} \tilde{F}_n^{-1/2} F_n^{1/2}) \\
&= \lambda_{\max}(\tilde{V}_n(\beta^*))^{-1} \\
&= 1 / \lambda_{\min} \tilde{V}_n(\beta^*) \\
&\leq (1 - W_n(C, \beta_0))^{-1},
\end{aligned}$$

denn für den Eigenvektor τ von $\lambda_{\min} \tilde{V}_n(\beta^*)$ mit $\tau' \tau = 1$ gilt

$$\begin{aligned}
\lambda_{\min} \tilde{V}_n(\beta^*) &= \tau' \tilde{V}_n(\beta^*) \tau = \tau' \tau + \tau' (\tilde{V}_n(\beta^*) - I) \tau, \\
&\geq 1 - W_n(C, \beta_0).
\end{aligned}$$

Hieraus folgt die Behauptung. \square

Als nächstes betrachten wir nun das Konvergenzverhalten der Zufallsgrößen

$$\lambda_{\min}(F_n(\hat{\beta}_n)) \quad \text{sowie} \quad \hat{W}_n(\delta) = W_n(\delta, \hat{\beta}_n)$$

unter $\beta_0 \in B$:

Lemma 3:

Unter (D) und (N) gilt:

$$(a) \quad \lambda_{\max}(F_n(\hat{\beta}_n))^{-1} \xrightarrow{P_{\beta_0}} 0 \text{ nach } P_{\beta_0}\text{-Wahrscheinlichkeit}$$

und

$$(b) \quad \hat{W}_n(\delta) \xrightarrow{P_{\beta_0}} 0 \text{ nach } P_{\beta_0}\text{-Wahrscheinlichkeit.}$$

für jedes $\delta > 0$.

Beweis: Wir führen zunächst den Nachweis von (a): Für beliebiges $c > 0$ und $\delta > 0$ impliziert (N) die Existenz eines $n_0 \in \mathbb{N}$ derart, daß

$$\begin{aligned} \tau' F_n(\beta) \tau &\geq \tau' F_n(\beta_0) \tau - \left| \tau' F_n(\beta) \tau - \tau' F_n(\beta_0) \tau \right| \\ &\geq (1 - c) \tau' F_n(\beta_0) \tau \end{aligned}$$

für $\beta \in N_n(\delta, \beta_0)$ (vgl. auch F&K, (3.3)). Wiederum folgt somit für beliebiges $\eta > 0$ die Existenz eines $\delta > 0$ und $n_3 > 0$ derart, daß

$$\begin{aligned} P(\lambda_{\max}(F_n(\hat{\beta}_n))^{-1} < \zeta) &= P(\lambda_{\min} F_n(\hat{\beta}_n) > \zeta^{-1}) \\ &= P(\inf_{\tau, \tau=1} \tau' F_n(\hat{\beta}_n) \tau > \zeta^{-1}) \end{aligned}$$

$$\leq P(\hat{\beta}_n \notin N_n(\delta, \beta_0)) + P(\inf_{\tau, \tau=1} \tau' F_n(\beta_0) \tau > ((1-c)\zeta)^{-1}) \\ \leq \eta + 0$$

Hieraus folgt (a).

Nachweis von (b): Für $C > 0$ wählen wir $n_0 \in \mathbb{N}$ zunächst so groß, daß $W_n(C, \beta_0) < 1$ für $n > n_0$; ferner sei $\delta > 0$ beliebig sowie ferner $n > n_0$ und $\hat{\beta}_n \in N_n(C, \beta_0)$ fest vorgegeben.

Dann gilt für $\delta_n = \delta (1 - W_n(C, \beta_0))^{-1/2} + C$ gemäß Lemma 2:

$$W_n(\delta, \hat{\beta}_n) = \sup_{0 \neq \tau \in \mathbb{R}^q} \sup_{\beta \in N_n(\delta, \hat{\beta}_n)} \left| \frac{\tau' F_n(\beta) \tau}{\tau' F_n(\hat{\beta}_n) \tau} - 1 \right| \\ \leq \sup_{0 \neq \tau \in \mathbb{R}^q} \sup_{\beta \in N_n(\delta_n, \beta_0)} \left| \frac{\tau' F_n(\beta) \tau}{\tau' F_n(\hat{\beta}_n) \tau} - 1 \right| \\ \leq \sup_{\tau} \sup_{\beta \in N_n(\delta_n, \beta_0)} \left| \frac{\tau' F_n(\beta) \tau}{\tau' F_n(\hat{\beta}_n) \tau} - 1 \right| \left| \frac{\tau' F_n(\hat{\beta}_n) \tau}{\tau' F_n(\beta_0) \tau} \right| \\ \times (1 - W_n(C, \beta_0))^{-1} \\ \leq \sup_{\tau} \sup_{\beta \in N_n(\delta_n, \beta_0)} \left| \frac{\tau' F_n(\beta) \tau}{\tau' F_n(\beta_0) \tau} - 1 + 1 - \frac{\tau' F_n(\hat{\beta}_n) \tau}{\tau' F_n(\beta_0) \tau} \right| \\ \times (1 - W_n(C, \beta_0))^{-1}$$

$$\leq \sup_{\tau} \sup_{\beta \in N_n(\delta_n, \beta_0)} \left[\left| \frac{\tau' F_n(\beta) \tau}{\tau' F_n(\beta_0) \tau} - 1 \right| + \left| 1 - \frac{\tau' F_n(\hat{\beta}_n) \tau}{\tau' F_n(\beta_0) \tau} \right| \right] \\ \times (1 - W_n(C, \beta_0))^{-1}$$

$$\leq (W_n(\delta_n, \beta_0) + W_n(C, \beta_0)) / (1 - W_n(C, \beta_0))$$

Zu vorgegebenen $\eta > 0$ existiert nun ein $C > 0$ und $n_1 > n_0$ derart, daß für $n > n_1$

$$P_{\beta_0}(\hat{\beta}_n \in N_n(C, \beta_0)) \geq 1 - \eta$$

Zu $v > 0$ gibt es nun ferner $n_2 > n_1$ mit

$$(W_n(\delta_n, \beta_0) + W_n(C, \beta_0)) / (1 - W_n(C, \beta_0)) < v \quad \text{für } n > n_2.$$

Damit ist

$$P_{\beta_0}(\hat{W}_n(\delta) \leq v) \geq P_{\beta_0}(\hat{\beta}_n \in N_n(C, \beta_0)) \geq 1 - \eta$$

für $n > n_2$. Hieraus folgt die Behauptung. \square

Nach diesen Vorbereitungen führen wir nun den

Beweis von Satz 2: Sei $\{m(n)\}$ eine beliebige, streng monotone Folge in \mathbb{N} . Es ist zu zeigen, daß $\{m(n)\}$ eine Teilfolge $\{k(n)\}$ besitzt, für die

$$D_{k(n)}^S(\hat{\beta}_{k(n)}) \rightarrow N(0, I) \quad P_{\beta_0} \text{ - fast sicher.}$$

Nach Lemma 3 existiert $\{k(n)\}$ so, daß

$$\hat{W}_{k(n)}(\delta) \rightarrow 0 \quad P_{\beta_0} \text{ - fast sicher.}$$

Für eine feste Realisation $\{\hat{\beta}_{k(n)}^* \} = \{\hat{\beta}_{k(n)}(\omega)\}$, sei nun $(\hat{W}_{k(n)}(\delta))(\omega) \rightarrow 0$ erfüllt. Dann ist aufgrund von Lemma 1 und der daran anschließenden Bemerkung offenbar auch

$$D_{k(n)}^S(\hat{\beta}_{k(n)}(\omega)) \rightarrow N(0, I)$$

erfüllt. Hieraus folgt die Behauptung. |

Das bisher verwendete Beweisverfahren braucht nun für den Nachweis von Satz 3 nur geringfügig modifiziert zu werden:

Beweis von Satz 3: Der erste Teil der Aussage ist trivial: Offenbar verändert sich die Aussage von (D) und (N) nicht, wenn man dort $F_n(\beta)$ durch $F_n(\beta, \phi_0)$ (bzw. $F_n(\beta_0)$ durch $F_n(\beta_0, \phi_0)$)

ersetzt. Danach lässt sich die Beweisführung von F&K problemlos übernehmen.

Auch beim Nachweis des zweiten Teils kann im wesentlichen unsere Beweisführung übertragen werden. Es ist jedoch zu beachten, daß nun Folgen $\{(\beta_n^*, \phi_n^*)\}$ zu berücksichtigen sind und deshalb die Aussagen $D(\tilde{B})$ und $(N(\tilde{B}))$ sich verändern, falls dort $F_n(\beta)$ (bzw. $F_n(\beta_n^*)$) durch $F_n(\beta, \phi_n^*)$ (bzw. $F_n(\beta_n^*, \phi_n^*)$) ersetzt wird, es sei denn, daß die Folge $\{a(\phi_n^*)\}$ einen (positiven) Grenzwert besitzt. Dann existiert zu jedem $\delta > 0$ ein $\tilde{\delta} > 0$ und ein $\tilde{\delta} > 0$ derart, daß für $n > n_0$ (in naheliegender Terminologie)

$$\begin{aligned} \beta \in N(\delta, \beta_n^*) &\longrightarrow \beta \in N(\delta, \beta_n^*, \phi_n^*) \text{ und} \\ \beta \in N(\delta, \beta_n^*, \phi_n^*) &\longrightarrow \beta \in N(\tilde{\delta}, \beta_n^*) \end{aligned}$$

Damit ergibt sich die analoge Version von Lemma 1.

Lemma 2 muß nun umformuliert werden: Es wird folgende Aussage benötigt:

Sei $\{\phi_n^*\}$ so daß $a(\phi_n^*) \longrightarrow a(\phi_0)$. Ferner sei $\beta^* \in N_n(C; \beta_0, \phi_0)$ und $\beta \in N_n(\delta, \beta^*, \phi_n^*)$ für $C > 0$, $\delta > 0$, $\beta \in B$, $\beta^* \in B$, $\beta_0 \in B$. Dann gilt

$$\beta \in N_n(\tilde{\delta}_n, \beta_0) \text{ für } \tilde{\delta} \geq \delta (a(\phi_n^*) \cdot (1 - W_n(C, \beta_0)) / a(\phi_0))^{-1/2} + C.$$

Der Beweis dieser Aussage erfolgt leicht durch die Faktorisierung von $F_n(\beta, \phi)$.

Die Übertragung von Lemma 3 und somit von Satz 2 ist nun möglich, weil nach P_{β_0, ϕ_0} -Wahrscheinlichkeit $a(\hat{\phi}_n)/a(\phi_0) \rightarrow 1$ gilt. Damit ist auch Satz 3 bewiesen. |

Literaturverzeichnis

Beran, R. (1982). Estimated sampling distributions: the bootstrap and competitors.

Ann Statist. 10, 212-225

Beran, R. (1984). Jackknife approximations to bootstrap estimates.

Ann. Statist. 12, 101-118.

Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap.

Ann. Statist. 9, 1196-1217.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife.

Ann. Statist. 7, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.*

SIAM, Philadelphia.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy.

Statistical science 1, 54-77.

Fahrmeir, L. und Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models.

Ann. Statist. 13, 342-368.

Haberman, S.J. (1977). Log-linear models and frequency tables with small expected counts.

Ann. Statist. 5, 1148-1169.

Haberman, S.J. (1978). *Analysis of Quantitative Data Vol. I*
Academic Press, New York.

Haberman, S.J. (1984). *Log-concave Likelihoods and Maximum Likelihood Estimation.*

Preprint.

Nelder, J.A. und Wedderburn, R.W.M. (1972). Generalized linear models.

J. Roy. Statist. Soc. Ser. A 135, 370-384.

Payne, C.D. (ed.) (1985). *The GLIM System. Release 3.77.*
NAG, Oxford.

Sauermann, W. (1986) *Bootstrap-Verfahren in log-linearen Modellen.*
Dissertation, Heidelberg.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap.
Ann. Statist. 9, 1187-1195.

Tab.1: Reaktion männl. Befragter
(vgl. Text) auf Item des 74'er und
75'er GSS, klassifiziert nach Bil-
dungsniveau (*years of education in*
Jahren)

| Bildungs- niveau | Anzahl Ablehnung | Anzahl Zustimmung |
|---------------------|---------------------|----------------------|
| 0 | 2 | 4 |
| 1 | 0 | 2 |
| 2 | 0 | 4 |
| 3 | 3 | 6 |
| 4 | 5 | 5 |
| 5 | 7 | 13 |
| 6 | 9 | 25 |
| 7 | 15 | 27 |
| 8 | 49 | 27 |
| 9 | 29 | 29 |
| 10 | 45 | 32 |
| 11 | 59 | 36 |
| 12 | 245 | 115 |
| 13 | 70 | 31 |
| 14 | 79 | 28 |
| 15 | 23 | 9 |
| 16 | 110 | 15 |
| 17 | 29 | 3 |
| 18 | 28 | 1 |
| 19 | 13 | 2 |
| 20 | 20 | 3 |

Tab.2: Unterstichprobe vom Umfang
100 aus Tab.1

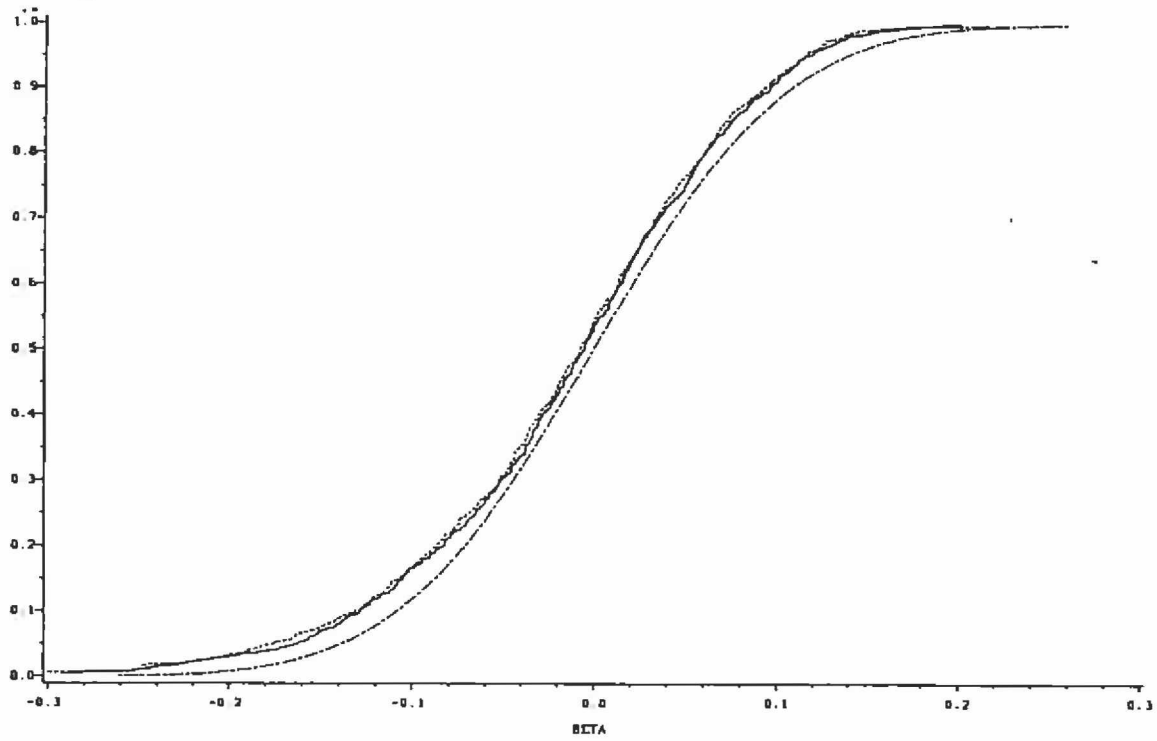
| Bildungs- niveau | Anzahl Ablehnung | Anzahl Zustimmung |
|---------------------|---------------------|----------------------|
|---------------------|---------------------|----------------------|

| | | |
|----|----|---|
| 0 | 0 | 1 |
| 4 | 0 | 1 |
| 5 | 0 | 3 |
| 6 | 3 | 7 |
| 7 | 4 | 1 |
| 8 | 3 | 1 |
| 9 | 5 | 2 |
| 10 | 21 | 9 |
| 11 | 8 | 4 |
| 12 | 4 | 5 |
| 13 | 2 | 0 |
| 14 | 5 | 1 |
| 15 | 5 | 0 |
| 16 | 2 | 0 |
| 17 | 1 | 1 |
| 18 | 1 | 0 |

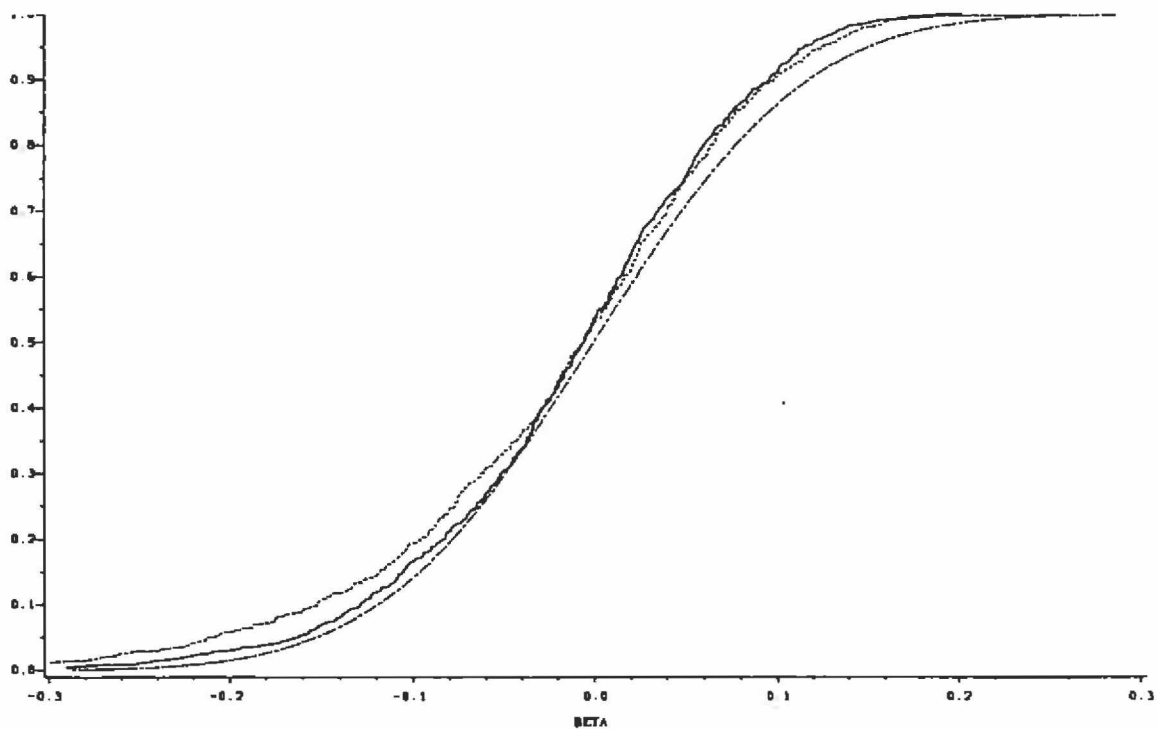
Fig.1:

Exakte Verteilung sowie Bootstrap- und Normalapproximation von $\hat{\beta}_1 - \beta_1$.

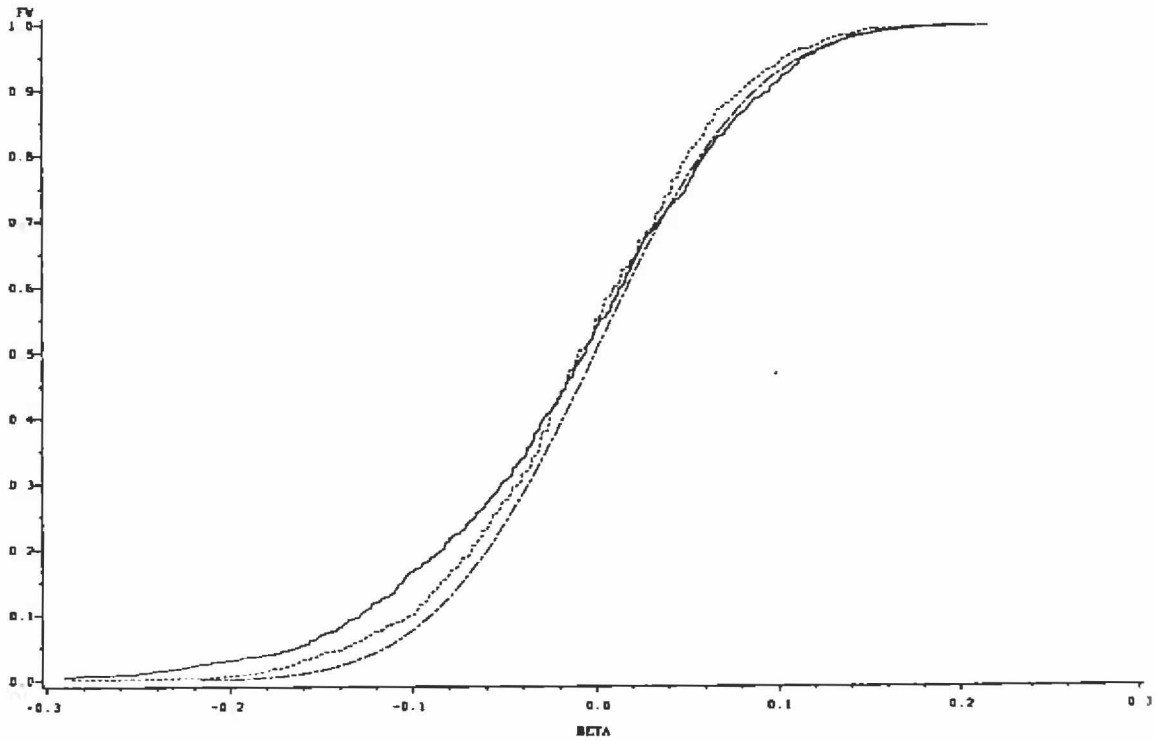
a) $\hat{\beta}_1 = -0.2757$



b) $\hat{\beta}_1 = -0.3313$



c) $\hat{\beta}_1 = -0.1278$



_____ exakte Verteilung
 - - - - - Bootstrap-Approximation
 - . - . - Normalapproximation

ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA-
Arbeitstagung vom 25.09. - 05.10.79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter
Schmidt
Methodenbericht zum "ALLBUS 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu
einer Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler
SAR - Simple AND Retrieval mit dem Siemens-EDT-Textmanipulations-
programm
- 82/14 Cornelia Krauth
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener
Two approaches to the analysis of judgments of prestige: Interindi-
vidual differences and the general scale
- 83/11 Rolf Porst
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-Fragen-
programms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und
ALLBUS 1982
- 84/01 Manfred Küchler, Peter Ph. Mohler
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen
Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten,
Teil I + II
- 84/02 Bernd Wegener
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-
Prestige-Skala

- 84/03 Peter Prüfer, Margrit Rexroth
Erfahrungen mit einer Technik zur Bewertung von Interviewerverhalten
- 84/04 Frank Faulbaum
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982
- 84/05 Jürgen Hoffmeyer-Zlotnik
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten
- 84/07 Gabriele Hippler, Hans-Jürgen Hippler
Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the-Face" Technique
- 85/01 Hartmut Esser
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von Antwortverzerrungen in Interviews
- 85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang
Methodenbericht zum "ALLBUS 1984"
- 86/01 Dagmar Krebs
Zur Konstruktion von Einstellungsskalen im interkulturellen Vergleich
- 86/02 Hartmut Esser
Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung
- 86/03 Bernd Wegener
Prestige and Status as Function of Unit Size
- 86/04 Frank Faulbaum
Very Soft Modeling: The Logical Specification and Analysis of Complex Process Explanations with Arbitrary Degrees of Underidentification and Variables of Arbitrary Aggregation and Measurement Levels
- 86/05 Peter Prüfer, Margrit Rexroth (Übersetzung: Dorothy Duncan)
On the Use of the Interaction Coding Technique
- 86/06 Hans-Peter Kirschner
Zur Kessler-Greenberg-Zerlegung der Varianz der Meßdifferenz zwischen zwei Meßzeitpunkten einer Panel-Befragung
- 86/07 Georg Erdmann
Ansätze zur Abbildung sozialer Systeme mittels nicht-linearer dynamischer Modelle
- 86/08 Michael Braun, Rolf Porst
ALLBUS-Bibliographie (5. Fassung, Stand: 30.06.86)
- 86/09 Heiner Ritter
Einige Ergebnisse von Vergleichstests zwischen den PC- und Mainframe-Versionen von SPSS und SAS

86/10 Hans-Peter Kirschner
Der Stichprobenplan zum Projekt ISSP 1985 und seine Realisierung